



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2020

Prudens quaestio in diagnostic studies

Papageorgiou, Spyridon N

DOI: <https://doi.org/10.1177/1465312520970946>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-194610>

Journal Article

Accepted Version

Originally published at:

Papageorgiou, Spyridon N (2020). Prudens quaestio in diagnostic studies. Journal of Orthodontics, 47(4):363-365.

DOI: <https://doi.org/10.1177/1465312520970946>

TITLE PAGE

Prudens quaestio in diagnostic studies

Spyridon N. Papageorgiou¹

¹ Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Zurich, Switzerland

Corresponding author: Spyridon N Papageorgiou, Clinic of Orthodontics and Pediatric Dentistry, Center of Dental Medicine, University of Zurich, Plattenstrasse 11, Zurich CH 8032, Switzerland.

Email: snpapage@gmail.com.

Words in text: 1092

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Spyridon N Papageorgiou <https://orcid.org/0000-0003-1968-3326>

BLINDED MANUSCRIPT

A fictional scenario.....

In the previous article (Papageorgiou, 2020) I discussed a fictional scenario about an orthodontist trying to retrospectively assess the prevalence of gingival recession among patients within his private practice using already existing documentation—with the main focus being on the effect of varying observation periods (treatment duration). I take the previous scenario as an initial basis and build upon it. The aim of this piece is to think about the method with which the event of interest (here gingival recession) is identified. Or in other words, what is the actual performance of a screening test?

Suppose that a group of 200 orthodontic patients drawn randomly from the past archives of a private practice, where in truth a total of 68 patients (34%) with at least one gingival recession associated with any tooth side after treatment. Suppose that this percentage of 34% reflects by omniscience or a near-perfect screening test ('gold standard') the actual proportion of patients with gingival recession. In order for the orthodontist to gain an estimate of the prevalence of gingival recession among his patients, a specific screening test has to be established and used on the patient sample. As is usual in medicine, several different screening tests might exist that all aim to identify if an outcome of interest exists. In this specific example, development of gingival recession might be ascertained, among others, through patient-reported observations, through evaluation of the patient's intraoral photographs, through evaluation of the patient's dental casts, through prospective clinical assessment by the orthodontist, or by comparison of intraoral scans.

For this theoretical scenario, we are not interested per se in the specific diagnostic performance of the various screening tests mentioned above, but rather we want to focus on aspects of screening performance in general. Therefore, the five different screening tests are named Test 1-5 (T1-T5) and their theoretical performance in identifying gingival recession is contrasted to the true percentage of 34%. Table 1 gives the comparison of T1 to the true prevalence.

In order to assess the diagnostic performance of T1, we need to calculate various measures, which include sensitivity, specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV).

Sensitivity tells us how often does a positive T1 finding pick up a recession or in other words, how sensitive is a test at picking up the disease. This is found by dividing the true positive by all the diseased patients (or dividing the patients with recession found by a positive T1 by all the patients actually having a recession), which is $55/68=80.9\%$.

Specificity tells us how often a negative T1 finding reflects an actual lack of recession or in other words, how specific is a test in isolating a disease. This is found by dividing the true negative by all healthy patients (or dividing patients where T1 indicate no recession by all patients actually not having a recession), which is $93/132=70.5\%$.

Positive Predictive Value (PPV) tells us how often a positive T1 actually finds an existing recession or in other words, how often is a patient with recession picked up by T1. This is found by dividing the true positive by all positive patients (or dividing patients having actually a recession as indicated by a positive T1 by all patients with a positive T1), which is $55/94=58.5\%$.

Finally, the Negative Predictive Value (NGV) tells us how often a negative T1 finds a recession-free patient or in other words, how often is a recession-free patient picked by T1. This is found by dividing the true negative by all negative patients (or dividing patients without actually a recession as indicated by a negative T1 by all patients with a negative T1), which is $93/106=87.7\%$.

Following the same thought process, the diagnostic performance of all five screening tests T1-T5 is given in Table 2.

Which of the following statements are correct if any?

- (A) If one wants to be sure that as many of the recession patients as possible are picked up by a positive test, then T4 outperforms the other tests.
- (B) If one wants to be sure that as many of the recession-free patients as possible are picked up by a negative test, then T2 outperforms the other tests.
- (C) In a hypothetical case of a deadly disease, for which a widely-available and relatively safe treatment exists, one would prefer T3 as a diagnostic modality.
- (D) In a hypothetical case of a condition, where a positive answer has an important impact on the patient's life, but a delayed diagnosis is not necessarily detrimental, one would prefer T5 as a diagnostic modality.

The ability of a positive test to correctly identify truly diseased (here, recession-ridden) persons or how 'sensitive' is a test towards a disease is reflected in the sensitivity. Test T4 has however the highest PPV, not sensitivity. The PPV tells if we get a positive test result, how sure we can be that the test was correct and the patient had indeed a recession. The correct answer therefore is that T3 performs best in terms of sensitivity and (A) is wrong.

The ability of a negative test to correctly rule out the existence of disease (here, the lack of recession) or how 'specific' is a test in isolating the disease is reflected in the specificity. T5 has actually the highest specificity and therefore (B) is wrong (even though if its sensitivity is 0%).

It is of course always desirable for a diagnostic test to have the maximum sensitivity and specificity so that both diseased and healthy patients can be most efficiently identified. In reality however, sensitivity and specificity have an inverse relation, so that when sensitivity increases, specificity decreases. And this is where critical appraisal on an individual basis for each clinical scenario comes.

Consider for example that we have a very deadly disease, for which we have a treatment and this treatment is easily tolerable. In that case, we much rather we don't miss any real cases of the disease. In that case, having a very high sensitivity in favour of impaired specificity might actually make sense. This will enable all patients that really need treatment to receive it as soon as possible. Inadvertently, some healthy patients that might get a false-positive test result might also end up receiving treatment, but as this treatment has been hypothesized to be relative harmless, this should not be any problem. Therefore, the most appropriate test for this is T3 and C is correct.

Take now another example, where a positive test result is a big deal and might have considerable consequences, whereas a delayed diagnosis is not that detrimental. Think about a pregnancy test, where

you are going to find out sooner or later, usually without any considerable consequences, and you don't want to have many people being upset by a false-positive test without reason. In this case, the test with the lowest rate for false-positive results might be preferable, which is T4, so answer (D) is false.

Sensitivity and specificity are well-known attributes evaluated in diagnostic accuracy studies, but that is usually not what the average (lay-)person wants to know. People want to know if they have a positive result, what percentage of people will have a disease. Or, conversely, if they have a negative result, what percentage of people will be healthy. This is understood by most people as "what the tests mean". Even though PPV and NPV are intuitively important characteristics that may supplement sensitivity / specificity, they also carry considerable drawbacks, which will be further evaluated in the next piece.

References

As time goes by (2020) As time goes by. J Orthod 47: 273–275.

Table 2. Comparative diagnostic performance of the five different diagnostic tests for identifying gingival recessions compared to the true prevalence of gingival recession among the included patients

| | | In truth | | |
|------------|-------------------|---------------|---------------|-------|
| | | Recession (+) | Recession (-) | Total |
| T1 results | Find recession | 55 | 39 | 94 |
| | Find no recession | 13 | 93 | 106 |
| Total | | 68 | 132 | |

T1, test 1.

Table 2. Comparative diagnostic performance of the five different diagnostic tests for identifying ingival recessions compared to the true prevalence of gingival recession among the included patients

| | T1 | T2 | T3 | T4 | T5 |
|---------------------|-----------|-----------|-----------|-----------|-----------|
| Sensitivity | 80.9% | 89.7% | 92.7% | 27.9% | 0.0% |
| Specificity | 70.5% | 79.6% | 54.6% | 97.0% | 100.0% |
| PPV | 58.5% | 69.3% | 51.2% | 82.6% | 34.3% |
| NGV | 87.7% | 93.8% | 93.5% | 72.3% | 66.0% |
| False negative rate | 12.3% | 6.3% | 6.5% | 27.7% | 33.7% |
| False positive rate | 41.5% | 30.7% | 48.8% | 17.4% | 65.9% |

T1-T5, test 1-5; NPV, Negative predictive value; PPV, Positive predictive value.